

The battle: grammatical complexity versus definiteness

A corpus-based study of PP placement in translated and non-translated Dutch

One of the variation phenomena where hearer and speaker motivations play a decisive role, is the extraposition vs. non-extraposition of a prepositional phrase (PP) in Dutch. Consider the following example, where both word order variants (1) *deelnamen aan de Olympische Spelen* (PP after V-final; extraposition) and (2) *aan de Olympische Spelen deelnamen* (PP before V-final; non-extraposition) are grammatically equivalent.

- | | |
|---|--|
| 1 | [...] dat de Belgen deelnamen V-final <u>aan de Olympische Spelen</u>
[...] that the Belgians take part in the Olympics |
| 2 | [...] dat de Belgen <u>aan de Olympische Spelen</u> deelnamen V-final
[...] that the Belgians in the Olympics take part |

Although this type of variation has received considerable attention until the early 1990's (e.g. Jansen 1978, Braecke 1990), it is still unclear what exactly drives the choice between PP extraposition vs. non-extraposition in Dutch. On the basis of a large-scale corpus of modern translated and non-translated Dutch and by means of multivariate statistical techniques, this paper aims to shed light on the variables that explain the variation in PP placement. More particularly, we focus on the effect of the variables, as previous research in Dutch and other languages has shown that they have an important impact on different types of syntactic variation (see, among others, Arnold et al. 2000, De Sutter 2007, van Bergen & de Swart 2010). The general assumption is that language users have a tendency to extrapose grammatically complex PPs (e.g. Hawkins 1994, Wasow 2002) and that indefinite PPs are extraposed more often compared to definite PPs (see Van Nieuwborg 1967). Furthermore, we want to establish whether the variables *grammatical weight* and *definiteness* have an identical effect in translated and non-translated texts, since previous research made it clear that the language use in translated and non-translated texts differ systematically on all linguistic levels (e.g. Olohan & Baker 2000). Nevertheless, most of the corpus-based studies on variation in non-translated and translated texts are restricted to comparing the general distribution of linguistic variants. The factors that determine this variation remain uninvestigated.

In order to reach these goals, we extracted all PPs from the Dutch Parallel Corpus (DPC), which is a 10-million-word, parallel corpus of Dutch, English and French (Macken et al. 2011). It contains six different text types, but for the present study we opted to look only at PP variation in journalistic texts. After manually checking the relevance of the extracted corpus data, we obtained a data set containing 1732 dependent clauses initiated with the grammatical conjunction *dat* (= *that*) and with a PP either before V-final or after V-final. These data were coded for grammatical complexity (in terms of number of words) and definiteness (definite vs. indefinite).

The results of the logistic regression analysis (table 1) show a.o. that not only the grammatical complexity of the PP, but also the grammatical complexity of the *middlefield* (all constituents before V-final; *MF*) and the *postfield* (all constituents after V-final; *PF*) have an effect on the word order variation. The assumption that complex PPs are placed more at the end of the sentence compared to short PPs can be confirmed. Further, we discovered that these variables have a similar effect in translated and non-translated texts (figure 1). The effect of the variable *definiteness* and its relationship to *grammatical weight* will be presented at the workshop.

	O.R	P-waarde
Translations from EN	-	p > .05
Translations from FR	1.4	P = 0.06
Length of the PP	0.72	P < .0001 ***
Length of the MF	0.93	P < 0.01 ***
Length of the PF	1.11	P < .0001 ***

Table 1

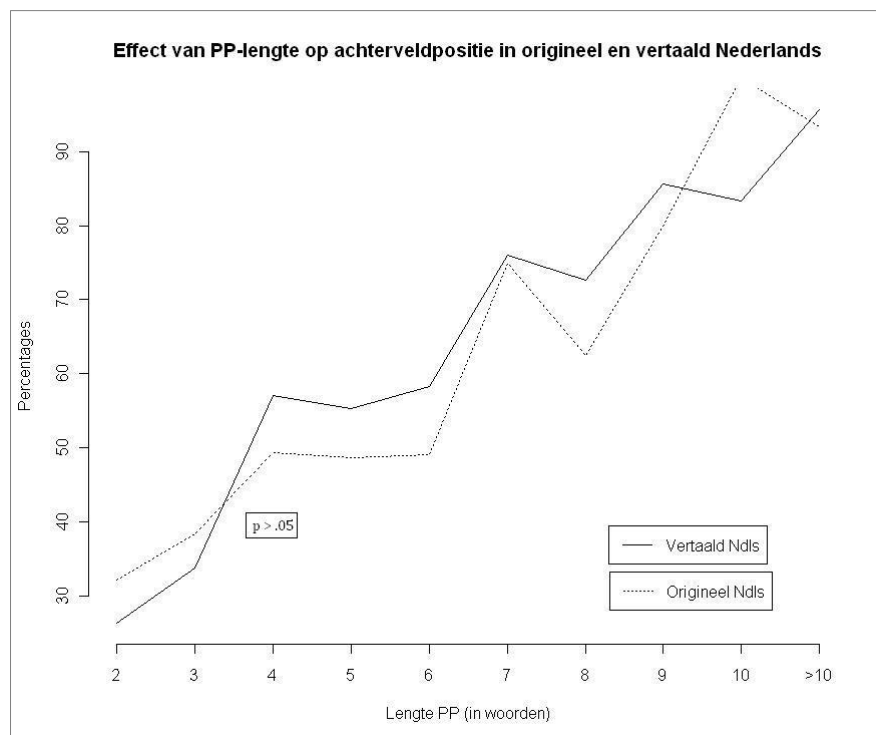


Figure 1

References:

- Arnold, J. E., T. Wasow, A. Losongco & R. Ginstrom (2000). Heavyness versus newness: The effects of structure complexity and discourse status on constituent ordering. *Language* 76 (1). 28-55
- Braecke, C. (1990). Uit de tang of ± prominent? *Taal en Tongval* 3 (Dialectsyntaxis), 125-134
- De Sutter, G. (2007). Naar een corpusgebaseerde, cognitief-functionele verklaring van de woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen. *Nederlandse Taalkunde* 12 (4). 302-330
- Jansen, F. (1978). Hoe krijgt een spreker zijn woorden op een rijtje? *Taalgebruikaspecten van de 'PP over V' constructie*. J. G. Kooij (ed.) *Aspecten van de woordvolgorde in het Nederlands*. Leiden: Vakgroep Nederlandse Taal- & Letterkunde. 70-104.

- Macken, L., O. Declercq & H. Paulussen (2011) Dutch Parallel Corpus: a Balanced Copyright-Cleared Parallel Corpus. *META* 56(2)
- Olohan M. & M. Baker (2000). Reporting that in translated English: evidence for subconscious processes of explicitation? *Across Languages and Cultures* 1(2). 141-158.
- Van Bergen, G & P. de Swart (2010). Scrambling in spoken Dutch: Definiteness versus weight as determinants of word order variation. *Corpus Linguistics and Linguistic Theory* 6 (2). 267-295.
- Nieuwborg, E. (1967). Vallen in het Nederlands de plaatsingscategorieën samen met de traditionele ontledingscategorieën? *Handelingen 27ste Vlaamse filologencongres*.
- Wasow, T. (2002). *Postverbal behavior*. Stanford, CA: CSLI Publications.